



SCIENCE & NATURE

Know Thy Futurist

CATHY O'NEIL

Image: Maurizio Pesce



curio.io
Boston Review | Know Thy Futurist

SOUNDCLOUD

Cookie policy



Have you heard? Someday we will live in a perfect society ruled by an omnipotent artificial intelligence, provably and utterly beneficial to mankind.

That is, if we don't all die once the machines gain consciousness, take over, and kill us.

Wait, actually, they are going to take some of us with them, and we will transcend to another plane of existence. Or at least clones of us will. Or at least clones of us that are not being perpetually tortured for our current sins.

Futurism is the American dream on overdrive: a disdain for the status quo and a belief that we can solve it all without unions, public education, and social safety nets.

These are all outcomes that futurists of various stripes currently believe. A futurist is a person who spends a serious amount of time—either paid or unpaid—forming theories about society’s future. And although it can be fun to mock them for their silly sounding and **overtly religious predictions**, we should take futurists seriously. Because at the heart of the futurism movement lies money, influence, political power, and access to the algorithms that increasingly rule our private, political, and professional lives.

Google, **IBM, Ford, and the Department of Defense** all employ futurists. And I am myself a futurist. But I have noticed deep divisions and disagreements within the field, which has led me, below, to chart the four basic “types” of futurists. My hope is that by better understanding the motivations and backgrounds of the people involved—however unscientifically—we can better prepare ourselves for the upcoming political struggle over whose narrative of the future we should fight for: tech oligarchs that want to own flying cars and live forever, or gig economy workers that want to someday have affordable health care.

With that in mind, let me introduce two dimensions of futurism, represented by axes. That is to say, two ways to measure and plot futurists on a graph, which we can then examine more closely.

The first measurement of a futurist is the extent to which he or she believes in a singularity. Broadly speaking a singularity is a moment where technology gets so much better, at such an exponentially increasing rate, that it achieves a fundamental and meaningful technological shift of existence, transcending its original purpose and even nature. In many singularity myths the computer either becomes self-aware and intelligent, possibly in a good way but sometimes in a destructive or even vindictive way. In others humans are connected to machines and together become something new. The larger point is that some futurists believe fervently in a singularity, while others do not.

On our second axis, let’s measure the extent to which a given futurist is worried when they theorize about the future. Are they excited or scared? Cautious or jubilant? The choices futurists make are often driven by their emotions. Utopianists generally focus on all the good that technology can do; they find hope in cool gadgets and the newest AI helpers. Dystopianists are by definition focused on the harm; they consequently think about different aspects of technology altogether. The kinds of technologies these two groups consider are nearly disjoint, and even where they do intersect, the futurists’ takes are diametrically opposed.

So, now that we have our two axes, we can build quadrants and consider the group of futurists in each one. Their differences shed light on what their values are, who their audiences are, and what product they are peddling.



Q1.

First up: the people who believe in the singularity and are not worried about it. They welcome it with open arms in the name of progress. Examples of people in this quadrant are Ray Kurzweil, the inventor and author of ***The Age of Spiritual Machines*** (1999); the libertarians in the **Seasteaders movement** who want to create autonomous floating cities outside of any government jurisdiction; and the people who are trying to augment intelligence and live forever.

These futurists enthusiastically believe in Moore’s Law—the observation by Gordon Moore, a co-founder of Intel, that the number of **transistors** in a circuit doubles approximately every two years—and in exponential growth of everything in sight. Singularity University, started by Kurzweil, has no fewer than twelve mentions of the word “exponential” on **its website**. Its motto is “Be Exponential.”

Futurists are ready to install hardware in their brains because, as young or middle-age white men, they have never been oppressed.

Generally speaking these futurists are hobbyists—they have the time for these theories because, in terms of wealth, they are already in the top 0.1 percent. They think of the future in large part as a way to invest their money and become even wealthier. They once worked at or still own Silicon Valley companies, venture capital firms, or hedge funds, and they learned to think of themselves as deeply clever—possibly even wise. They wax eloquent about meritocracy over expensive wine or their drug of choice (**micro-dosing**, anyone?).

With enormous riches and very few worldly concerns, these futurists focus their endeavors on the only things that could actually threaten them: death and disease.

They talk publicly about augmenting intelligence through robotic assistance or better quality of life through medical breakthroughs, but privately they are interested in technical fixes to physical problems and are impatient with the medical establishment for being too cautious and insufficiently innovative. They invest heavily in cryogenics, dubious mind–computer interface technology, medical strategies for living forever (here’s **looking at you**, Sergey Brin and Larry Page), and possibly even the blood of young people.

These futurists are ready and willing to install hardware in their brains because, as they are mostly

young or middle-age white men, they have never been oppressed. For them the worst-case scenario is that they live their future lives as uploaded software in the cloud, a place where they can control the excellent virtual reality graphics. (If this sounds like a science fiction fantasy for sex-starved teenagers, don't be surprised. They got most of these ideas—as sex-starved teenagers—from writers such as Robert Heinlein and Ayn Rand.)

The problem here, of course, is the “I win” blind spot—the belief that if this system works for me, then it must be a good system. These futurists think that racism, sexism, classism, and politics are problems to be solved by technology. If they had their way, they would be asked to program the next government. They would keep it proprietary, of course, to keep the hoi polloi from gaming the system.

And herein lies the problem: whether it is the nature of existence in the super-rich bubble, or something distinctly modern and computer-oriented, futurism of this flavor is inherently elitist, genius-obsessed, and dismissive of larger society.



Q2.

Next: people who believe in a singularity but are worried about the future. They do not see the singularity as a necessarily positive force. These are the men—majority men, although more women than in the previous group—who read dystopian science fiction in their youth and think about all the things that could go wrong once the machines become self-aware, which has a small (but positive!) probability of happening. They spend time trying to estimate that probability.

A community center for these folks is the website lesswrong.com, which was created by Eliezer Yudkowsky, an artificial intelligence researcher. Yudkowsky **thinks** people should use rationality and avoid biases in order to lead better lives. It was a good idea, as far as practical philosophies go, but eventually he and his followers got caught up in increasingly abstract probability calculations using Bayes' Theorem and bizarre thought experiments.

There is no need to imagine some hypothetical future of human suffering. We are already here.

My favorite is called **Roko's basilisk**, the thought experiment in which a future superintelligent and powerful AI tortures anyone who imagined its existence but didn't go to the trouble of creating it. In other words it is a vindictive hypothetical being that puts you in danger as soon as you hear the

thought experiment. Roko's basilisk was seen by its inventor, Roko, as an incentive to donate to the cause of Friendly AI to "thereby increase the chances of a positive singularity." But discussion of it soon so dominated Yudkowsky's site that he banned it—a move that, not surprisingly, **created more interest** in the discussion.

A different but related movement in the world of AI futures comes from **the Effective Altruism movement**, which has been **advocated for in this journal** by philosopher **Peter Singer**. Like Yudkowsky, Effective Altruists started out well. Their basic argument was that we should care about human suffering outside our borders, not just in our close proximity, and that we should take personal responsibility for optimizing our money to improve the world.

You can go pretty far with that reasoning—and to their credit, Effective Altruists have made enormous international charitable contributions—but obsessing over the concept of effectiveness is limited by the fact that suffering, like community good, is hard to quantify.

Instead of acknowledging the limits of hard numbers, however, the group has more recently spun off into a parody of itself. Some factions believe that instead of worrying about current suffering, they should worry about "existential risks," unlikely futuristic events that are characterized by computations besieged by powers of ten and could thus cause enormous suffering. A good example comes from **Nick Bostrom's Future of Humanity Institute website**: ". . . we find that the expected value of reducing existential risk by a mere one billionth of one billionth of one percentage point is worth a hundred billion times as much as a billion human lives."

As a group these futurists are fundamentally sympathetic figures but woefully simplistic regarding current human problems. If they are not worried about existential risk, they are worried about **the suffering of plankton**, or perhaps every particle in the universe.

I will shove Elon Musk into this Q2 group, even though he is not a perfect fit. Being an enormously rich and powerful entrepreneur, he probably belongs in the first group, but he sometimes shows up at Effective Altruism events, and he has made noise recently about the **computers getting mean and launching us into World War III**. The cynics among us might suspect this is mostly a ploy to sell his services as a mediator between the superintelligent AI and humans when the time inevitably comes. After all Musk always has something to sell, including **a ticket to Mars**, Earth's backup planet.



Q3.

Next: the technoutopianists. They don't specifically talk about the singularity, but they are willing to extol the many virtues of technology to people who want to believe them.

They latch on to the latest idea—e.g., will Bitcoin solve the world's problems?—and turn it into a paid speech. They are not super wealthy, but they aspire to be wealthier and more famous. Follow the money here and you will find that they are what *The Ideas Industry* (2017) author Daniel Drezner would call “thought leaders,” single-idea merchants paid by oligarchs to feel special at TED or TED-like conferences.

These are the New Prophets of Capital, **as spelled out by Nicole Aschoff**, and they will peddle whatever depoliticized fad captures the attention of the super rich at a given time. With Steve Jobs as their patron saint, they represent the American dream on overdrive—a disdain for the status quo and the notion that we can solve it all without the old, outdated trappings of unions, public education, and social safety nets. They have no time for taking on difficult questions of structural inequality that do not fade away with the wave of a magical wand.

Most futurists are talking about sci-fi fantasies. Why?

Far from actually fixing problems, they are the type of futurist that is most obviously selling something: a corporate vision, blind faith in the titans of industry, and the sense of well-deserved success. **Alida Draudt**, for example, is a futurist employed by Capital One, the credit card company. At a recent Lesbian Who Tech conference, she explained to the audience that her job was to focus on “positive futures” so that Capital One could make those futures a reality. It is not entirely clear what that means, but I doubt it means free credit for everyone.

There are more women still in this group, but in the end that does not denote progress. This is the slick and ingratiating sales force for the futurism movement. Their aim is to control the conversation and, in repeating predictions about the future often enough, to cause that future to become a fixed, normalized idea in our collective imagination—even if that means a surveillance state with good shopping.



Q4.

Finally, the people who do not believe in singularities, but who are still worried. This is my group. In the fourth quadrant, we have no money and not much influence. We are majority women, gay men, and people of color. We are underrepresented at the data science institutes popping up all over the country because the commercial goals of such places are inconsistent with our inconvenient cries of

concern.

And I am concerned. Because from **personality tests that filter out** qualified job applicants to **crime risk algorithms** that convince judges to issue longer sentences, automated algorithms are already replacing our most important human decision making processes. As I look around, I realize there is no need to imagine some hypothetical future of human suffering. We are already here. Data scientists are creating machines they do not fully understand, machines that separates winners from losers for reasons that are already very familiar to us: class, race, age, disability status, quality of education, and other demographic measures. It is a threat to the very concept of social mobility. It is the end of the American dream. And yet most futurists are talking about sci-fi fantasies. Why?

At the heart of futurism lies money, influence, political power, and access to the algorithms that increasingly rule our private, political, and professional lives.

In a recent public talk, Yann LeCun, the director of AI at Facebook, was careful to distinguish between AI and algorithms. The study of the game Go used to be considered AI, he explained, but now we have a machine-learning algorithm that beats even the best human Go player. This is a sly trick of the hand; it means we are never talking about what is already happening. But what is already happening is by no means trivial. Games aside, the Facebook algorithm is already sufficiently powerful to manipulate our democracy. It doesn't matter if the people who work in AI don't want to consider it as AI.

Moreover, the Q1 technologists and the Q3 technoutopianists always refer to either chess or Go when painting their pretty picture of the future because those are two instances where everyone can agree on what success looks like: there is a clear winner and a clear loser. But that clarity of purpose and model of success gets a lot trickier on a planet that is quickly losing its supply of clean water and burning too much fossil fuel. In a hypothetical world where people could live forever—gobbling up resources indefinitely and exerting political influence with outdated political frameworks—should we allow them to?

In the end my taxonomy (as amusing as I find it) doesn't really matter to the average person. For the average person there is no difference between the singularity as imagined by futurists in Q1 or Q2 and a world in which they are already consistently and secretly shunted to the “loser” side of each automated decision. For the average person, it doesn't really matter if the decision to keep them in wage slavery is made by a super-intelligent AI or the **not-so-intelligent Starbucks Scheduling System**. The algorithms that already **charge people with low FICO scores more for insurance**, or send black people to prison for longer, or send more police to already over-policed

neighborhoods, with facial recognition cameras at every corner—all of these look like old fashioned power to the person who is being judged.

Ultimately this is all about power and influence. The worst-case scenario is not a vindictive AI or Sergey Brin not getting to celebrate his two-hundredth birthday. In the worst-case scenario, e-capitalism continues to run its course with ever-enlarging tools at its disposal and not a skeptical member of the elite in sight.

While we have you...

...we need your help. You might have noticed the absence of paywalls at *Boston Review*. We are committed to staying free for all our readers. Now we are going one step further to become completely ad-free. This means you will always be able to read us without roadblocks or barriers to entry. It also means that we count on you, our readers, for support. **If you like what you read here, help us keep it free for everyone by making a donation.** No amount is too small. You will be helping us cultivate a public sphere that honors pluralism of thought for a diverse and discerning public.

SUPPORT BOSTON REVIEW

More In Science & Nature

